第九章-HDR 质量评价技术

本章聚焦 HDR 质量评价技术，对于编解码、色调映射以及逆色调映射等不同任务，通常会采取不同的评价方法。本章将从主观评价和客观评价两个角度对常用的 HDR 视觉质量评价技术做整体介绍。

9.1 HDR 主观评价方法

国际电信联盟组织 ITU 根据不同使用场景给出不同的 HDR 主观视频质量测试方案。本文主要参考 ITU-R Rec. BT.500[1]和 Rec. ITU-R BT.1788[2]建议书中的测试方法，列出几种具有代表性的、在实际研究应用过程中使用频率较高的测试方法：单刺激法（ITU-R Rec. BT.500）、双刺激连续质量标度方法（ITU-R Rec. BT.500）、SAMVIQ 方法（ITU-R Rec. BT.1788）；并对不同的 EOTF 曲线的评估质量和 CSM 测试模式做了介绍。

9.1.1 单刺激法 SSM

在单刺激法（Single Stimulus Methods）中，测试素材只包含测试序列，不包含相应的原始序列。测试前，会使用模拟演示序列为每位测试者讲解测试流程。测试时，对每一位测试者采用不同的随机顺序。测试者则对观看的每一个视频分别打分。测试流程如下图 1 所示。

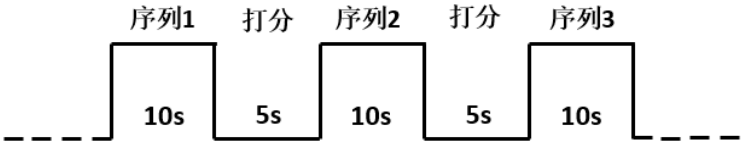

图 9-1 单刺激法测试流程

单刺激法通常采用五级质量量表，如下表 1 所示。该量表中分值为 MOS（Mean Opinion Score）分值，表示是对“视频质量”进行打分，即，认为该序列质量越高，则分数越高。有时，该方法也会采用九级或十一级质量量表，这两种情况时，即为最高分值为 9 分或 11 分。

表 1 单刺激法质量和损伤量表

MOS 分值	描述	损伤
5	优	不可察觉
4	良	可察觉，但不讨厌
3	中	稍微讨厌
2	差	讨厌
1	劣	很讨厌

该种测试方法的优点是首先较为符合测试者的实际日常使用场景，即，通常人们在看视频时，并不会事先观看无损源参考视频。其次，这种方法最大化了测试者每分钟可观看视频数量，节约了测试成本。最后，在严谨地进行测试设计和指导演示的情况下，即使是由不同批次的测试者进行测试，它依然具有稳定的可重复性。单刺激法的缺点也源自于它不与无损源参考视频比较这一优点，当需要检测某些特定细微差别如颜色变化的视频时则无法用此种方法进行测试。

9.1.2 双刺激连续质量标度法 DSCQS

在双刺激连续质量标度方法（Double Stimulus Continuous Quality Scale）中，需要对每个测试图像的两种状态 A&B 进行评分。其中一个是来自信号源的图像，即基准图像 A；另一个可能是经过被测系统输出的图像，即被测图像 B。基准和被测图像交替显示两次或多次之后（通常是两次）进行评分。

不同测试图像的一连串显示评分过程中，基准和被测图像呈现的先后次序以伪随机方式变动(观看员事先并不知道哪一个是基准图像)，要求观看员只简单地对每对图像的总体质量进行评分，并在评分表上作出标记。测试流程见下图 2。

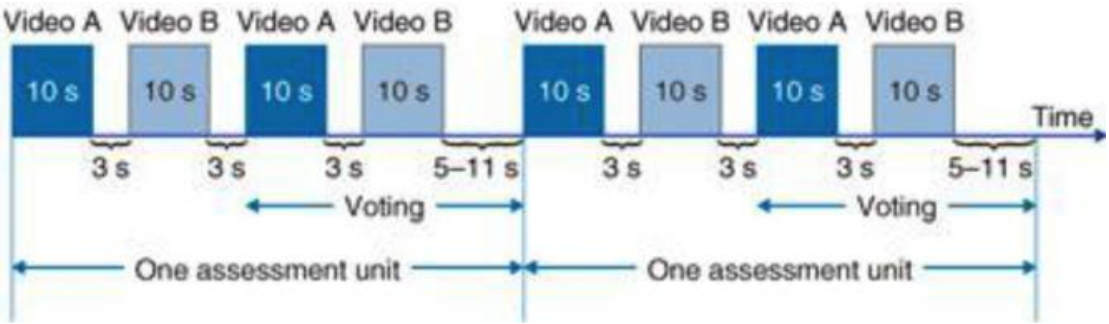

图 9-2 双刺激连续质量标度法测试流程

评分表由若干对纵向标度线组成，以适应对每个测试图像两种状态的评分。与单刺激法为避免量化误差，标度线提供连续标度,且被分成 5 个等级,相当于标准的 5 级质量标度范围。一个典型的 DSCQS 评分量表见下图 3。

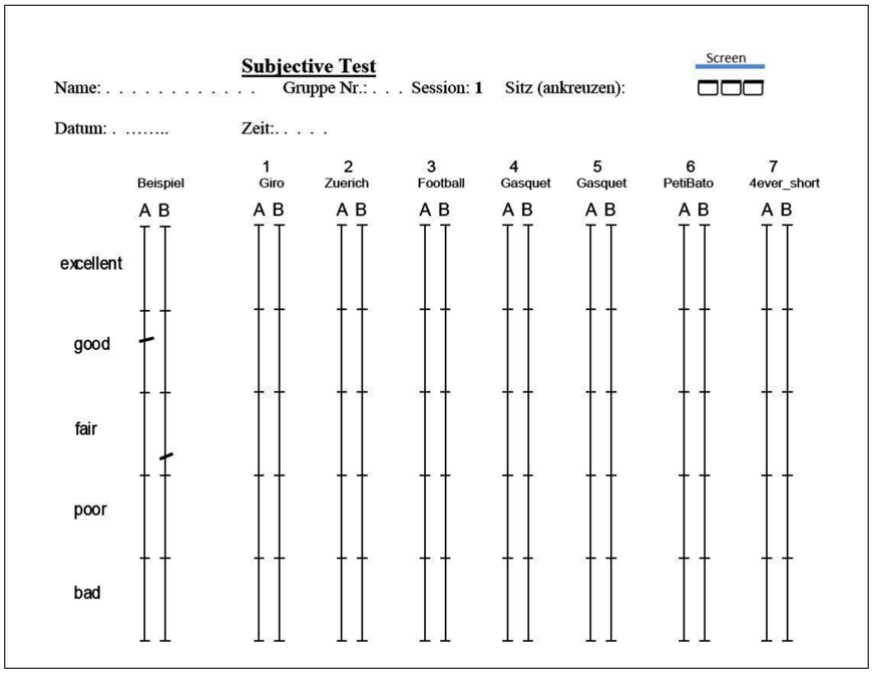

图 9-3 双刺激连续质量标度法测试样表

9.1.3 SAMVIQ 方法

在 SAMVIQ（Subjective Assessment Method for Video Image Quality）方法中，观测者准许观看一个片段的若干个版本。当所有版本都经观测者评定后，可对之后的片段内容进行评估。不同版本可由观测者通过计算机图形接口随机选择。根据需要，观测者可以停止、评审并修改某个片段各个版本的评分，也可以反复观看。

SAMVIQ 质量评估方法使用标度，以提供对视频片段质量的精准测量。各个观测者在从 0 到 100 评分的连续标尺上移动滑条，该连续标尺有 5 个线性排列的质量描述（优、良、中、差、劣）。其用户界面见下图 4。

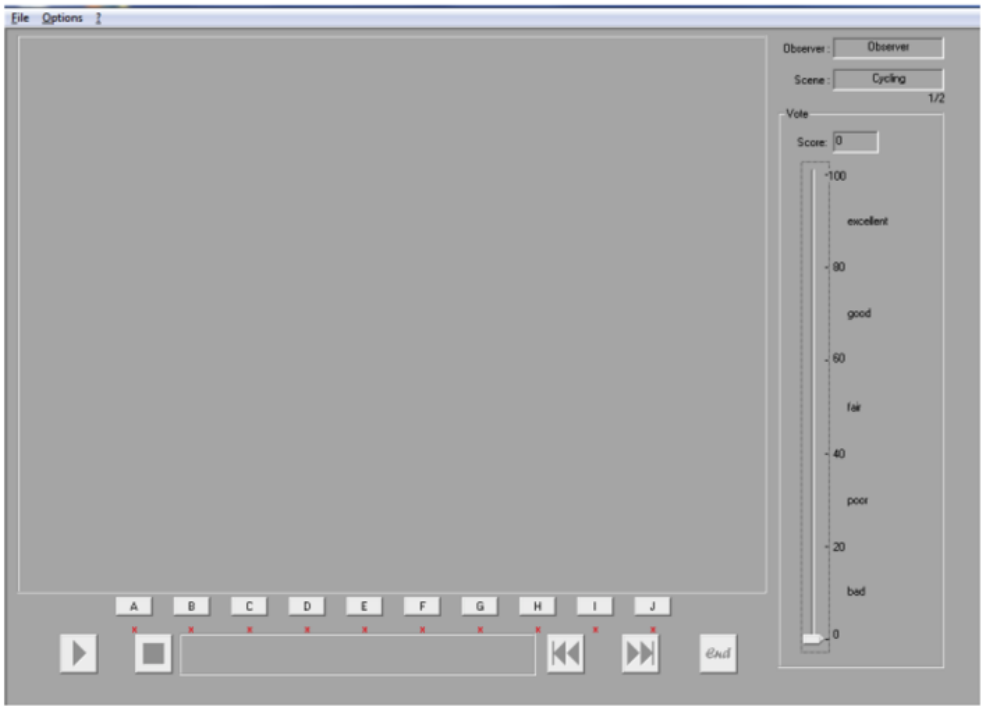

图 9-4 SAMVIQ 方法程序用户界面

SAMVIQ 包括一个显性基准（即未经处理的）原始片段，其评分通常为 100；以及相同片段的若干个版本，这些版本包括经处理的和未经处理的（即隐含基准）片段。。

经测试表明，可以使用显性基准来最大限度地缩小分值的标准差，尤其是对多媒体数字信号编解码器性能的评估。为了评估基准的内在质量，也可加上隐含基准分。值得注意的是，当没有可用的基准时，测试仍有可能进行，但标准的偏差会显著增大。

SAMVIQ 测试流程

a）逐个视频地进行测试。

b）对当前场景，可能以任何次序来播放任何片段，并为其打分。每个片段都可以多次播放和打分。

c）从一个场景到另一个场景，对片段的访问是随机的，防止观测者试图根据已排好的次序、以完全相同的方式来做出判定。实际上，在一个测试中，算法的次序仍保持相同，以便简化对结果的分析和陈述。只有来自相同按钮的相应访问是随机的。

d）对第一次观测，当前的片段必须在打分之前全部播放过；否则，可能立即打分和停止。

e）为测试下一个场景，必须为当前场景的所有片段打分。

f）为完成测试，必须为所有场景的所有片段打分。

典型的测试方式见下图 5。

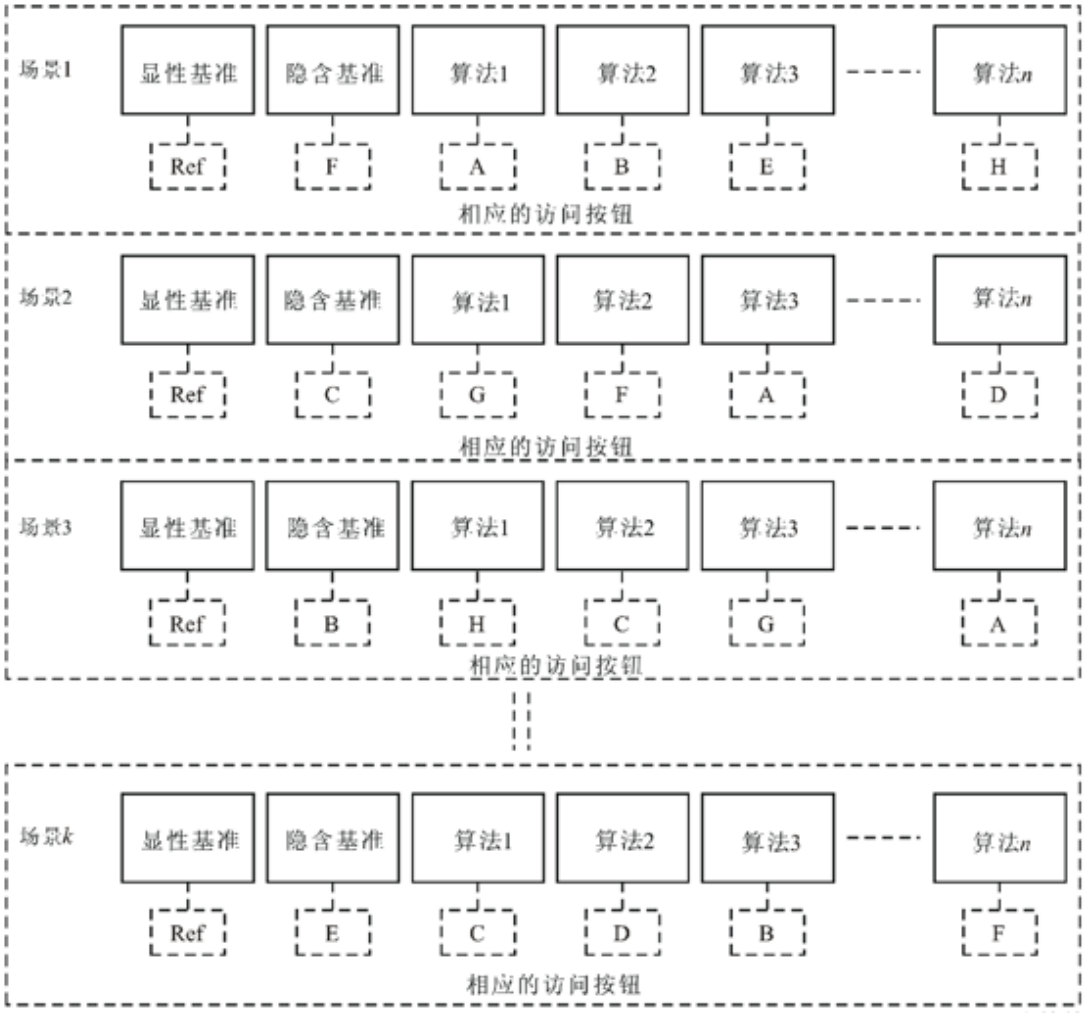

图 9-5 SAMVIQ 方法测试流程[2]

SAMVIQ 方法适用于多媒体内容，原因是它可能结合图像处理的不同特点，例如多媒体数字信号编解码器类型、图像格式、比特率、图像缩放等。

9.1.4 对于不同 EOTF 的主观评测

HDR 技术中常用的光电转换函数曲线有 PQ 曲线和 HLG（Hybrid Log Gamma 曲线）对于同样条件下拍摄和编码解码的 HDR 图像就产生了差别。

EBU 报告中采取 DSCQS 方法和 SAMVIQ 方法对于两种方案的质量进行了评测，其采用了不同比特率、不同编码方式的 HDR 视频进行测试，其结果见下图 6、图 7。

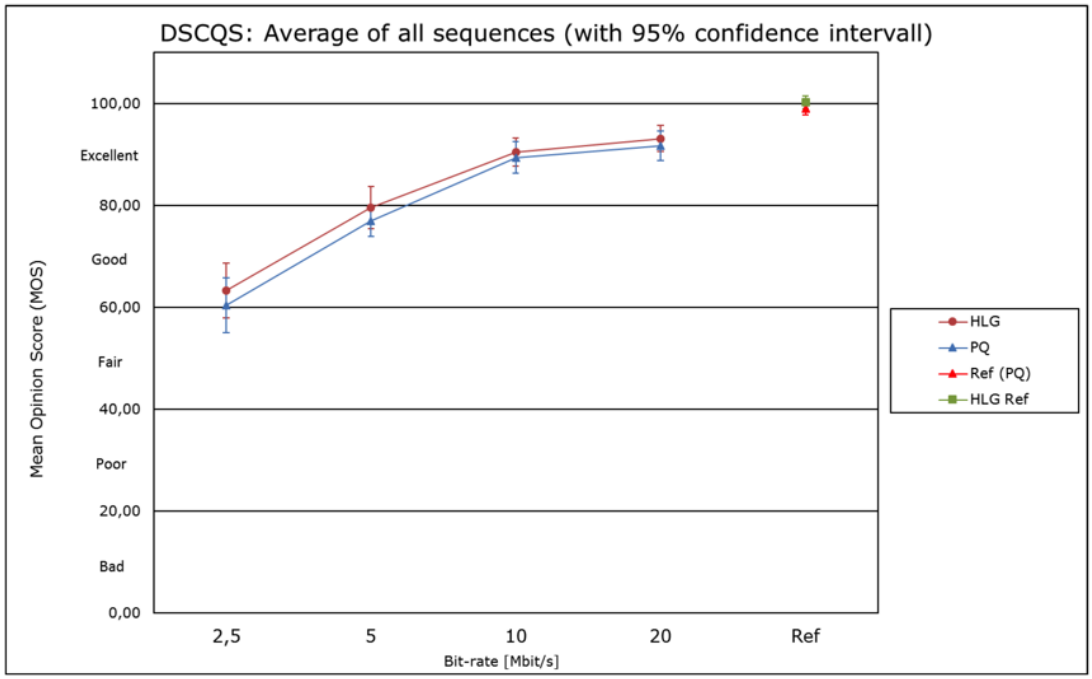

图 9-6 DSCQS 方法测试 PQ&HLG

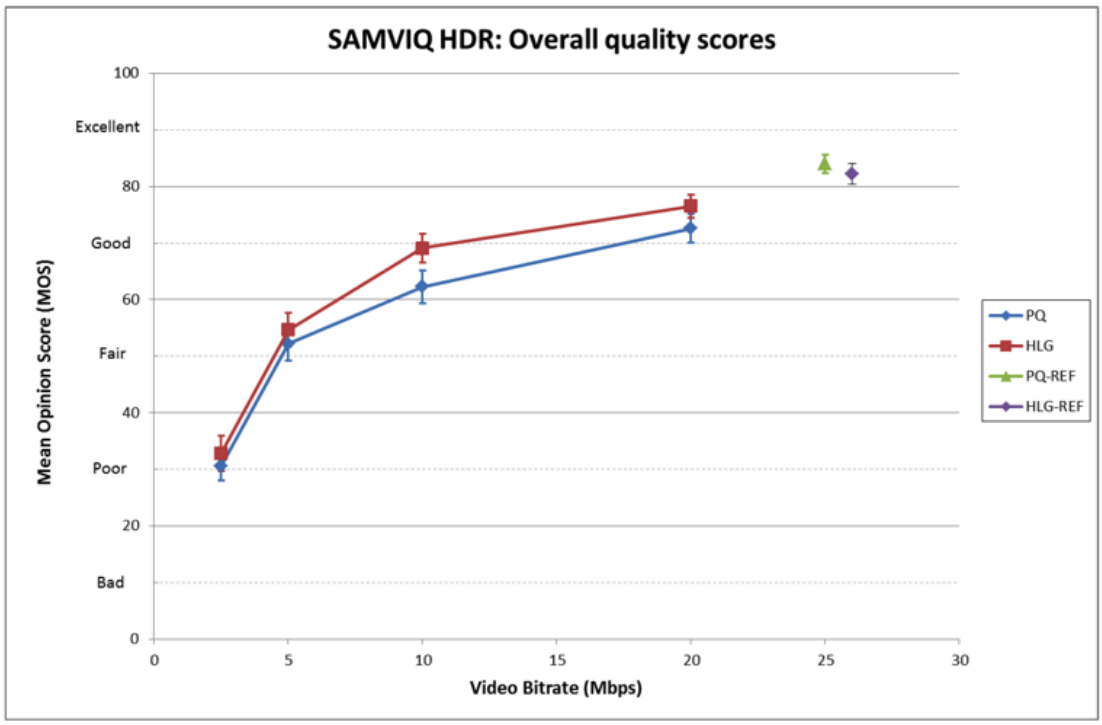

图 9-7 SAMVIQ 方法测试 PQ & HLG

由此可见，DSCQS 下的两种方法的主观评测质量几近相同，而 SAMVIQ 下 HLG 整体稍高于 PQ。两种方法在高码率下显示出高的 MOS 分。

9.1.5 Color Space and Monitor（CSM）测试模式

HDR 图像相较于 SDR 具有更高的亮度和色度，因此对比 SDR 只需要对于屏幕图像做直接的观察，HDR 的质量评价需要对颜色空间、显示器、亮度标准做更多的模式设计。SMPTE 设立的 CSM 测试模式[3]包含了十几种针对性的测试模式。下面将介绍常用的两种模式。

多空间：

通过展示八个颜色空间的同一特征图像，通过呈现效果可以辨别出其来源于哪种颜色空间，如下图即属于 BT.709;并且可以做到颜色空间的转换，如下图将 PQ2020 转换到 BT.709 并在 BT.709 的显示器上播放，则可以正确显示。

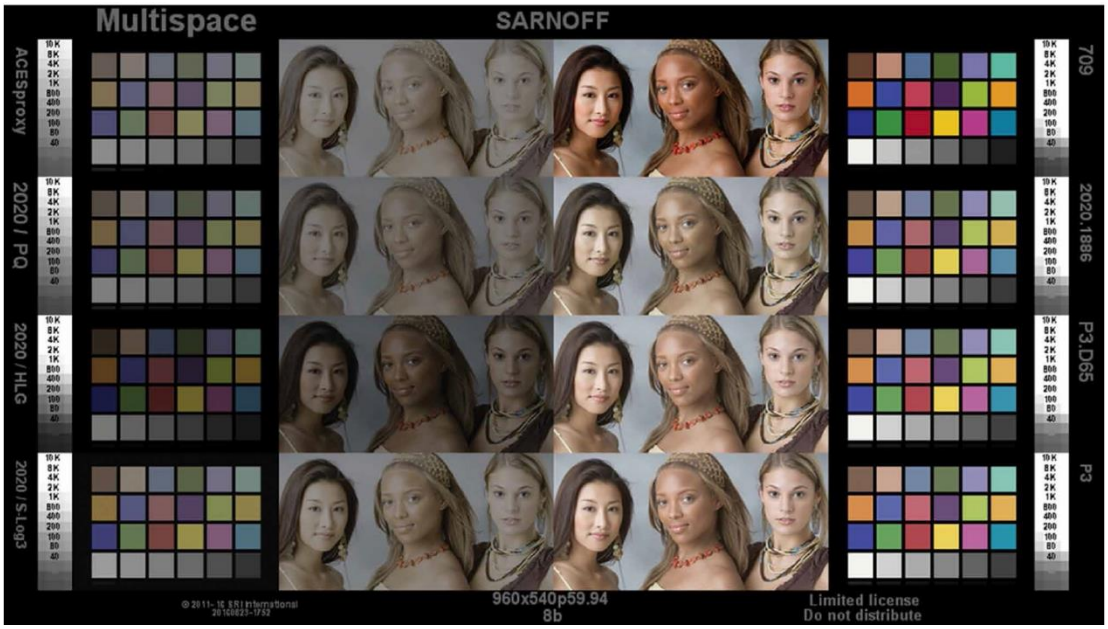

图 9-8 多空间模式[3]

三色标：

不同的颜色空间又不同的色标，因此可以通过把常用的 PQ2020、PQP3D65、BT.709 的色标组合到同一颜色容器（PQ2020）中，就可以直接用矢量转换做颜色空间的转换。见下图 9。

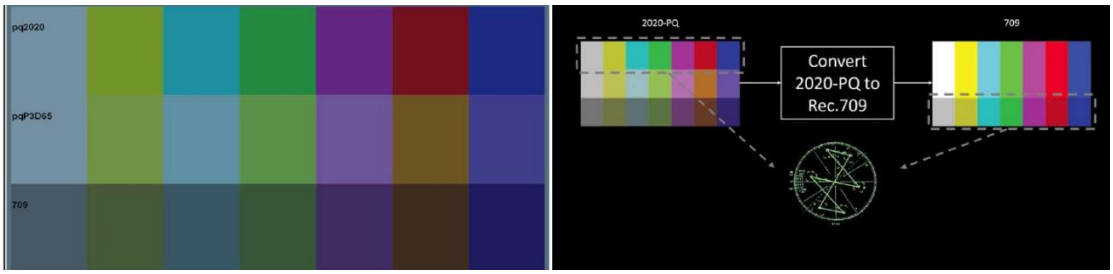

图 9-9 颜色空间的转换[3]

其他的模式简单介绍见下图 10。

TABLE 1. Some of the patterns included in the Color Space and Monitor (CSM) Suite.	
	**Multispace:** Faces, Macbeth Chart and Stair step rendered in eight color spaces in one image. Default spaces: ACESproxy, PQ/2020, HLG/2020, S-Log3, 709, 2020/1886, P3/D65, and P3. Custom spaces available.
	**Jacob's Ladder:** 103 calibrated steps covering 8 decades of brightness as 13 steps per decade. Shows tone mapping and clipping. Each decade strip contains a rising step to help verify the darkest visible level.
	**Color Ramps:** 28 rotating gradients in 7 colors covering 0-15 nits, 15-60, 60-240, and 240-960 nits. Gray background contains noise to challenge encoders. Reveals quantization banding issues of monitors and systems.
	**Code Values (Dark):** 10-bit labeled neutral patches covering code values 0 to 127. Allows determination of "Full" or "Limited/SMPTE" range interpretation of code values. Clearly shows groups of four patches when truncated to 8-bits.
	**Dark Moons:** (contrast enhanced here for printing). Rotating shallow gradients each starting at 0 nits and rising to a specified number of stops below 100 nits. Reveals how well monitors come out of black.
	**Dark Chips:** 10 blinking chips on a black background. Indicates the smallest step from black that can be seen on a given monitor in a given viewing environment.
	**Color Chips:** ST 303M "Macbeth" chip chart. Also includes Triple Gamut strip of saturated colors in three color spaces, and luma sweep. Use to verify correctness of color space conversion or interpretation.
	**Nit Chips:** A set of a dozen small-area brightness references. Useful for quick characterization of monitor brightness and tone mapping.
	**Zone Plate:** Two-dimensional frequency sweep of luma values. Reveals image scaling and resolution conversions.
	**Pixel Strips:** alternating rows of black and white lines (black/white/black/white) and colored lines (red/green/red/green) reveal whether a monitor's color resolution is as good as its black and white resolution. Also checks gamma matching and Y'CC matrix matching.
	**Triple Color Bars:** Bars from 2020, P3, and 709 in a 2020 container. Allows verification of color space conversion using a vectorscope.
	**Triple Gamut:** 18 color samples around the edge of each of three gamuts. With on-screen annotations of xyY values for each color. Scaled to 100 nits to avoid blown out colors.
	**Inter-Gamut:** Top row is colors around the 2020 gamut and bottom row is colors around the 709 gamut. Intermediate colors are evenly spaced (in x, y) between color pairs.
	**Full Gamut:** Top row is colors around the 2020 gamut and bottom row is neutral D65. Intermediate colors are evenly spaced in x, y between each 2020 color and D65.

图 9-10 CSM 模式列表[3]

9.1.6 主观评测方法总结

HDR 图像具有高亮度和宽色域的特点，因此其主观评测一般采取双刺激法，对不同的片段进行多次评估可以提高其准确性，但因此需要耗费大量的人力和测试成本。工业界提出多种客观评价方法，下一节将对其作详细介绍。

9.2 HDR 客观评价方法

HDR 客观质量评价方法与主观方法区别在于不需要观察者做主观评分，而是根据图像特征做直接的数据处理，具有更高的效率，根据失真图像其对于参考图像数据的需求程度可以分为三类：

全参考 FR（Full Reference）：将失真视频图像和无失真的源参考图像进行对比，评价失真图像相对源参考图像的质量损失

半参考 RR（Reduced reference）：从失真视频图像和无失真原始图像中分别提取图像层的某些有效特征，得出对失真视频帧的质量评价（不常用）

无参考 NR（No Reference）：不参考源图像

下面对 FR 方法和 NR 方法进行详细的分类。

9.2.1 全参考质量评价方法

根据[4]传统的全参考质量评价方法可分为以下几类：

面向统计的度量：

主要通过计算像素点值的不同来比较参考图像和失真图像的差距，常用的有 MSE 和 PSNR。但其对于 HDR 适用度较差，改进方案有：

1）mPSNR[5]通过对于不同曝光度独立计算 MSE 值再取平均；

2）PU-PSNR 通过在计算 MSE 前先做 Perceptually Uniform (PU)编码[6]从而考虑到了人类视觉系统对于亮度的非线性反应，从而更准确地评估了 HDR 图像的质量。

结构相似性度量：

通过比较亮度、对比度、结构等差异性来衡量失真性，常用的有 UQI[7]、SSIM[8]、MS-SSIM[9]、M-SVD[10]、QILV[11]等，PU-SSIM 也类似 PU-PSNR 做了编码来提高适用性。

视觉信息度量：

通过测量 HVS 的一些视觉特征或者视觉保真度来衡量失真性，常用的方法有 IFC[12]、VIF[13]、VIFp[13]、FSIM[14]等

信息权重度量：

通过对不同的区域设置局部权重代表其对失真的感知程度，来更符合感知标准地衡量失真性，常见的有 IW-MSE[15]、IW-PSNR[15]、IW-SSIM[15]等。

基于 HVS 的度量：

尝试模拟人类对于自然场景的感知，常见地有 JND_st[16]、WSNR[17]、DN[18]等。

颜色差异度量：

尝试补偿 CIE1976 的颜色梯度和感知的颜色的非线性，常用的有 CIE1976[19]、

CIE94[20]、CMC[21]、CIEDE2000[22]等。

限于篇幅，下面详细介绍两种最常用、准确性最高的适用于 HDR 图像的质量评价算法 HDR-VDP-2[23]以及 HDR-VQM[24]。

HDR-VDP-2

HDR-VDP-2[23]是一个用于比较参考图片和测试图片的视觉度量，并提供可见性和质量两个方面的预测信息：可见度——参考图片和测试图片之间的差异性能够被普通的观看者看得到的可能性；质量——测试图像相对于参考图片的质量退化，以平均意见得分 MOS 进行表达。

下图是 HDR-VDP-2 的质量评测标准的使用流程。其输入分别是测试图片和有质量损失的参考图片（一般是两个 HDR 图像或者两个 LDR 图像），然后经过 HDR-VDP-2 处理后产生一个检测概率图：整个概率值使用 0-1 的 P 表示，P 越大表示检测概率越大；质量预测使用 0-100 的参数 Q_{MOS}，Q 值越大表示质量越好。检测概率图告诉我们有多大的可能会感受到两个图的差异性，红色表示可能性较高，绿色表示可能性较低。由于失真是噪声和模糊的共同造成的，因此在平滑区域（噪声）和高对比度区域有最大的可能性检测到失真。

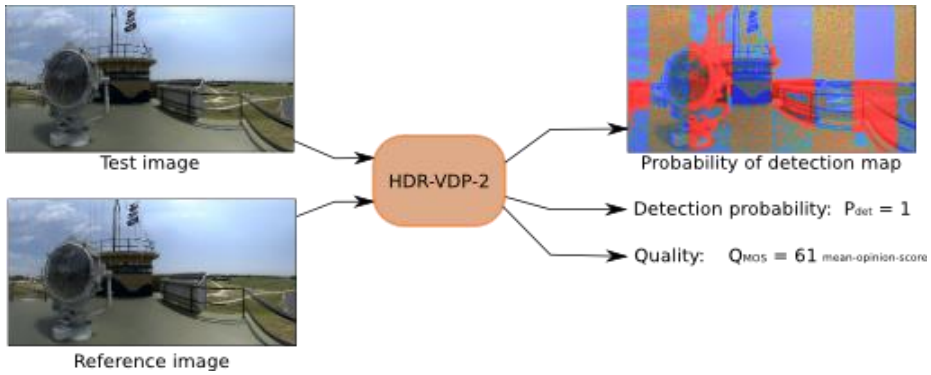

图 9-11　HDR-VDP-2 的输入及输出

尽管视觉差异性度量有很多，但相比于其他可见差异性度量，HDR-VDP-2 度量有其独特特点。首先，它可以应用到真实世界全范围亮度，即可以进行 HDR 的质量评价；其次，它对可见性和质量进行分别预测，这两个标准适用于不同目的且不相关；HDR-VDP-2 经过了严格测试和校准用于保证高精确度；最后，该方法代码开源：

http://hdrvdp.sourceforge.net/wiki/

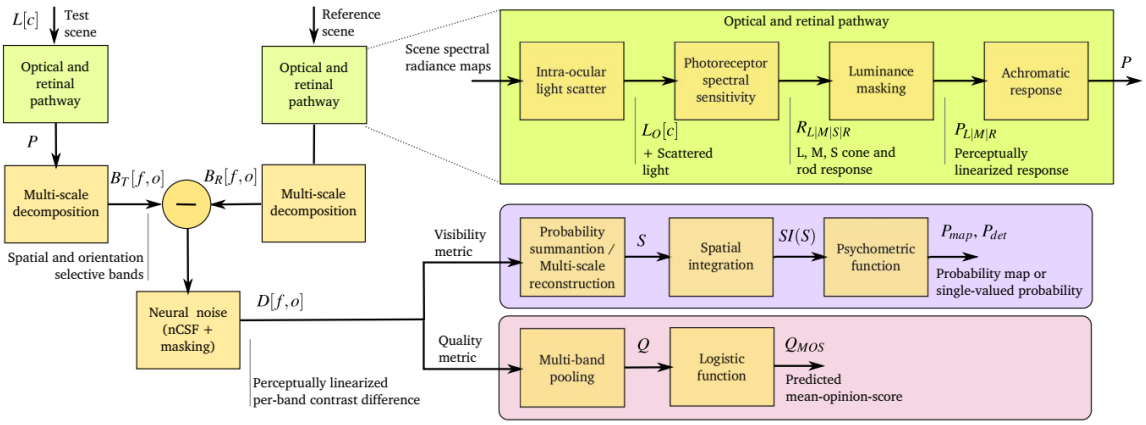

图 9-12 HDR-VDP-2 算法流程框图（包含可见性预测与质量预测）

HDR-VDP-2 主要包含的模块如图 12 所示，下面对各模块的主要功能和关键公式做简要介绍。

光学及视网膜通道模型（Optical and retinal pathway）

1）眼内光分散（Intra-ocular light scatter）

透过眼睛的光少部分分散在角膜、晶状体及视网膜上，这钟分散放大了图像高频信息的同时减弱了投射到视网膜上光的对比度。这种现象在观看含有强光 HDR 场景时更明显。光色散通过如下调制转移函数（Modulation Transfer Function, MTF）进行建模，该模型的输入是光谱图 L[c]：

$$F\{L_O\} = F\{L\}[c] \cdot \mathrm{MTF} \quad \text{(1)Equation Section (Next)Equation Section (Next)}$$

其中 F{·}表示傅里叶变换, [·]表示输入光谱图的标号。MTF 函数如下所示，其中 p 表示图像域以度为单位的频率周期。

$$\mathrm{MTF} = \sum_{k=1..4} a_k e^{-b_k p} \quad (2)$$

2）感光光谱灵敏度（Photoreceptor spectral sensitivity）

感光光谱灵敏度曲线表示光感受细胞可以感知特定波长光的概率，如下图所示对于 S/M/L 锥细胞及视杆细胞的敏感度曲线。

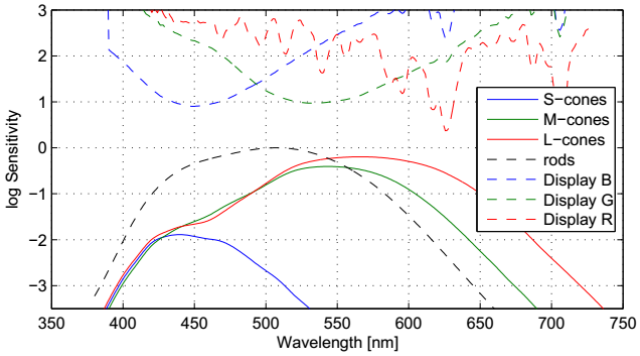

图 9-13 视锥及视杆细胞的感光光谱灵敏度曲线

当观测到具有频谱函数 f[c]的光时，不同感光细胞可以感知光的比例表示为

$$v_{L|M|S|R}[c]=\int_{\lambda}\sigma_{L|M|S|R}(\lambda)\cdot f[c](\lambda)d\lambda \quad (3)$$

其中σ表示视锥及视杆细胞的频谱敏感度，c 代表散射函数为 f[c]的输入频谱图标号。当给定 N 个输入频谱图时，每种感光细胞感受的总光量为

$$R_{L|M|S|R}=\sum_{c=1}^{N}L_O[c]\cdot v_{L|M|S|R}[c] \quad (4)$$

3）亮度掩盖（Luminance masking）

感光细胞对光波长敏感，对光强呈现高度非线性响应。人眼可观察很宽的物理光亮度范围，主要是由于感光细胞具有增益控制能力，可以根据接收光强调整感光敏感度。这种过程通常称为亮度掩盖。使用如下非线性感光函数 tL|M|R 描述上述调整机制。

$$P_{L|M|R}=t_{L|M|R}(R_{L|M|R}) \quad (5)$$

上述函数未考虑 S 锥细胞（因为此类细胞不承担亮度感知功能）。非线性感光函数如下：

$$t_{L|M|R}(r)=s_{peak}\int_{r_{min}}^{r}\frac{1}{\Delta r_{L|M|R}}d\mu=s_{peak}\int_{r_{min}}^{r}\frac{s_{L|M|R}(\mu)}{\mu}d\mu \quad (6)$$

其中 r 表示感光细胞的接收光（RL|M|R），rmin 是最小的可感知光强(10-6 cd/m2)，$\Delta r(r)$是检测阈值，SL|M|R 是感光细胞光敏感度，Speak 是调整视觉系统的敏感度峰值（需要针对不同的数据集调整）。SL|M|R 通过如下方式计算，首先计算不同感光细胞敏感度的和：

$$s_A(l)=s_L(r_L)+s_M(r_M)+s_R(r_R) \quad (7)$$

其中 l 是适应光强度（Adapting Luminance, l=La=RL+RM），sA 包含在 CSF 函数中，通过在每个亮度级别下最大化对比度敏感度求得

$$s_A(l)=\max_{\rho}\left(\mathrm{CSF}(\rho,\ l)\right) \quad (8)$$

由于不存在可以单独分析 L-和 M-锥细胞光敏感度的数据，因此假设 SL=SM，但是对于一些色盲患者的测定可以得到 R-杆细胞的光敏感度 SR，因此视锥细胞敏感度函数可以视为正常人与色盲患者的光敏感度函数之差决定：

$$s_{L|M}(r)=0.5(s_A(2r)-s_R(2r)) \quad (9)$$

对比度敏感函数（Contrast Sensitivity Fucntion, CSF）函数模型如下：

$$\mathrm{CSF}(\rho)=p_4 s_A(l)\frac{\mathrm{MTF}(\rho)}{\sqrt{\left(1+(p_1\cdot\rho)^{p_2}\right)\cdot\left(1-e^{-(\rho/7)^2}\right)^{-p_3}}} \quad (10)$$

其中ρ是以周期每角度（cycles-per-degree）为单位的空间域频率, p1..4 是在为不同的适应亮度 l 下的进行参数拟合得到的参数，其他亮度值对应参数通过使用对数亮度值作为插值系数对测定亮度值下的对比度函数值进行插值得到。sA(l)是视锥与视杆细胞的联合亮度敏感度（定义如式 7），通过如下计算模型得到：

$$s_A(l) = p_5\left(\left(\frac{p_6}{l}\right)^{p_7} + 1\right)^{-p_8} \quad (11)$$

通过调整 p4 及 p5 参数使得 CSF 与 sA(l)的比值峰值为 1.故 sA(l)可以计算得到。

4）消色差相应（Achromatic response）

为了计算视锥及视杆细胞的消色差响应，将它们对应相加：

$$P = P_L + P_M + P_R \quad (12)$$

相等权重是基于 L/M 细胞对于亮度感知的贡献相同。而视杆细胞对亮度的感知贡献 PR 由非线性感光函数 tL|M|R 控制，因此不需要额外的权重。

多尺度分解（Multi-scale decomposition）

一些研究表明视觉中枢中存在一种机制，该机制对于图像域频率和方向的某个特定范围具有选择性。为了模拟这种机制，视觉模型通常使用多尺度图像分解，一般利用小波变换或金字塔结构。本模型中使用了方向金字塔结构以对不同的图像域频率和方向进行分隔。与其他视觉分解方式相同，每个频率带的带宽随着频率的减少而减半。每幅图像分解为四个方向带以及该图像分辨率对应的最大空间域频率带数目。

神经噪声（Neural noise）

HDR-VDP 将对比度检测中的差异归于不同来源的噪声，影响不同频率带中的对比度检测的总体噪声等于信号独立噪声（即 neural contrast sensitivity function）以及信号依赖噪声（即 visual masking）。若参考图片和待检测图片的第 f 个频率带及第 o 个方向金字塔表示为 BT|R[f,o]，噪声归一化的信号差异为

$$D[f,o] = \frac{|B_T[f,o] - B_R[f,o]|^p}{\sqrt{N_{nCSF}^{2p}[f,o] + N_{mask}^2[f,o]}} \quad (13)$$

指数经试验测定 p 取 3.5，控制掩盖函数的形状。

1）神经对比度敏感函数（Neural contrast sensitivity function）

信号独立噪声可以通过测定 CSF 的实验测定。然而，为了使用 CSF 函数，需要抵消其中已经被考虑在 MTF 函数（公式 2）的光流信息以及被考虑在光感受函数内的亮度独立信息, 因此神经对比度敏感函数通过人眼光学的 MTF 函数与联合亮度敏感度函数 SA（公式 8）对 CSF 归一化得到的，又因噪声幅度与敏感度幅度成反比，因此噪声幅度为

$$N_{nCSF}[f,o] = \frac{1}{nCSF[f,o]} = \frac{\mathrm{MTF}(\rho, L_a) s_A(L_a)}{\mathrm{CSF}(\rho, L_a)} \quad (14)$$

ρ是空间域频率带 f 的峰值敏感度，通过如下式计算

$$\rho = \frac{n_{ppd}}{2^f} \quad (15)$$

其中 nppd 是输入图像给定没视角像素的数目，对于最高频率带 f=1.

2）对比掩盖（Contrast masking）

信号独立分量 Nmask 主要刻画对比度掩盖现象，该现象使得不均匀背景中的细微差别可见性降低。如果一个模式叠加到另一个具有类似空间域频率及方向的模式之上，

则前者的可见性降低，该现象被称为对比掩盖现象，通过如下三元素求和进行建模。

$$N_{mask}[f,o] = \frac{k_{self}}{n_f}\left(n_f B_M[f,o]\right)^q + \frac{k_{xo}}{n_f}\left(n_f \sum_{i=O\setminus\{o\}} B_M[f,i]\right)^q + \frac{k_{xn}}{n_f}\left(n_{f+1}B_M[f+1,o] + n_{f-1}B_M[f-1,o]\right)^q$$

(16)

其中第一行对应于自掩盖，第二行对应于跨方向掩盖，第三行对应于两个临近频率带的掩盖。kself, kxo，和 kxn 是控制不同来源掩盖的权重。第二行的 O 表示方向集合，指数 q 控制掩盖函数的斜率。nf=2-(f-1)用来使用 q 指数化前归一化项，BM[f, o]是频带 f 及方向 o 上的活动

$$B_M[f,o] = min\{|B_T[f,o]|,|B_R[f,o]|\}nCSF[f,o]$$

(17)

可见度指标（Visibility metric）

1）心理测量函数（Psychometric function）

公式（13）中将每个频带中的信号进行了归一化，D =1 对应于某个特定频率和方向选择机制的检测阈值。D[f,o] 通过如下心理测量函数映射为概率值 P

$$P[f,o] = 1-\exp(\log(0.5)D^{\beta}[f,o])$$

(18)

其中是测量函数的斜率，取值为 1，引入常数项 log(0.5)的目的是使得 D =1 时 P = 0.5。

2）概率集成（Probability summation）

Pmap 对应于不同频率带及方向的概率集成，表现为一张空间图，图中的每个像素点表示

$$\begin{aligned} P_{map} &= 1-\prod_{(f,o)}(1-P[f,o]) \\ &= 1-\prod_{(f,o)}\exp(\log(0.5)D^{\beta}[f,o]) \\ &= 1-\exp\left(\log(0.5)\sum_{(f,o)}D^{\beta}[f,o]\right) \end{aligned} \quad (19)$$

通过将乘积运算变为加法运算，可以使用可控金字塔（steerable pyramid）的重建变换将不同频带的概率相加，这样的重建变换包含每个频带内的重建滤波、上采样及将不同频带所得结果求和，与（19）中对不同频带间差异 D 的求和等价，因此在实际操作中使用可操控金字塔重建函数F^{-1}进行如下操作，SI 表示空间域集成函数:

$$P_{map} = 1-\exp\left(\log(0.5)SI\left(F^{-1}(D^{\beta})\right)\right) \quad (20)$$

3）空间域集成（Spatial integration）

使用空间域集成可以使得大型模式更容易被检测到，其数学表达如下

$$SI(S) = \frac{\sum S}{\max(S)}\cdot S \quad (21)$$

其中 S = $F^{-1}(D^{\beta})$表示对比度差异图。

质量指标（Quality metric）

一般情况下，重要的是视觉差异对图像质量的影响而非视觉差异本身。

HDR-VDP-2 设计目的主要用于预测视觉差异而不是质量，但是经过如下方式可转换为质量分数。

1）池化策略（Pooling strategy）

质量指标的主要目的是将一对待对比图片的进行感知层面上的区分，因此损伤的幅度值对应于视觉可见性而不是像素值之间的数学差值。HDR-VDP-2 通过对每一频带计算 D[f, o]实现这一目标，然而此值对应于不同频带内的像素值差异，需要将不同频带内的信息聚类为预测图像质量的单一数值。

为了找到最佳的聚类策略，作者在 LIVE 及 TID2008 数据库上（由于 HDR-DVP-2 操作在物理亮度值上，因此各数据库中的图片首先被转换为三色 XYZ 值，假设标准 LCD 显示器，CCFL 背光，sRGB 色域坐标，gamma = 2.2， 180 cd/m2 峰值亮度及 1 cd/m2 的黑值）测试了 20 种不同聚类方法，例如：最大值聚类，百分比聚类，以及 Minkowski 聚类，最终选择了如下聚类方式：

$$Q = \frac{1}{F \cdot O}\sum_{f=1}^{F}\sum_{o=1}^{O} w_f \log\left(\frac{1}{I}\sum_{i=1}^{I} D^2[f,o](i) + \varepsilon\right) \quad (22)$$

其中 i 表示像素下标，ε 表示避免当 D 值为 0 时运算结果溢出添加的常量(10-5), I 表示像素总数，wf 表示逐频带权重，初始值为 1，在 LIVE 数据库上使用褪火法进行优化。

2）逻辑回归映射（Logistics function）

客观预测值不是直接对应主观 MOS 值，需要如下的 logistic 非线性函数映射为主观分值：

$$Q_{MOS} = \frac{100}{1 + exp(q_1(Q + q_2))} \quad (23)$$

上述聚类以及映射过程使用 LIVE 数据库进行，使用 TID2008 数据库进行检验。

需要注意的是，目前 HDR-VDP-2 只进行亮度值的比较，而忽略色度信息，不对色度差异性进行检测。另外，它只能检测到参考图片和测试图片的差异性，给出基于差异的有参质量得分。

HDR-VQM

HDR-VQM[24]质量评价方法具有动态范围独立的特征，并且考虑到了时域变化通过 HVS 的时域模型，由于人类视觉系统倾向于在特定时间内关注特定区域，从而可在一个时空界内分析质量变化。其算法流程图见下图 14。

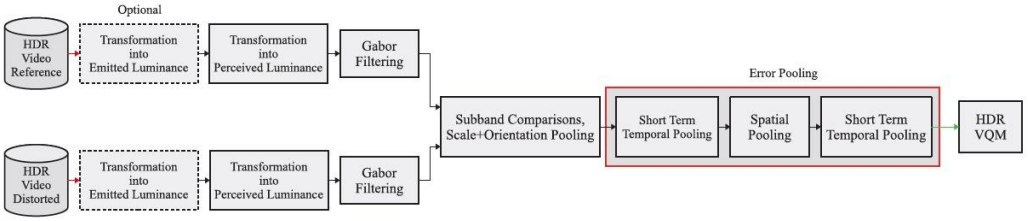

图 9-14 HDR-VQM 算法流程图[24]

第一阶段将信号的范围转换到给定的显示范围。这与 HDR 一样重要，因为不同的 HDR 显示具有不同的峰值亮度和动态范围。若源信号已参考给定范围，则不需要这一操作。

第二阶段，根据人类视觉系统对亮度的反应，利用 PU 编码将信号转换为基于感知的表示，从而保持视觉感知上具有均匀的间距。

子带误差阶段，在参考信号和失真信号之间产生了空间误差，并使用 log-Gabor 滤波器在不同尺度和方向上找到不同。子带对滤波后的信号取逆 DFT 来获得。其空间误差计算公式为：

$$Err_{t,s,o} = \frac{2l_{t,s,o}^{src} l_{t,s,o}^{dis} + \varepsilon}{(l_{t,s,o}^{src})^2 + (l_{t,s,o}^{dis})^2 + \varepsilon} \tag{24}$$

其中 $l_{t,s,o}^{src}$ 和 $l_{t,s,o}^{dis}$ 分别指参考视频和对应失真视频在第 t 帧，尺度 s，方向 o 上的子带，ε 是个小的正数，来避免非连续性。 则每帧的误差公式为：

$$Err_t = \frac{1}{N_{scale} \times N_{ori}} \sum_{1}^{N_{scale}} \sum_{1}^{N_{ori}} Err_{t,s,o} \tag{25}$$

考虑到短时记忆效应，最后阶段在临近非重叠时空区域将误差池化，并计算整体值其公式为：

$$HDR-VQM = \frac{1}{\left| t_s \in L_p \right|} \times \left| v \in L_p \right| \sum_{t_s \in L_p} \sum_{v \in L_p} ST_{u,t_s} \tag{26}$$

其中 L_p 表示最低 p%值的集合，$v \& t_s$ 分别空间和时间的下标，ST_{v,t_s} 为对应的误差帧。

FR 质量评价总结

Francesco 等人对于各类 FR 方法做了评测[25]，其评测标准是与主观测试评分做相似度估计，常用的标准吻合程度通常用 LCC 线性相关系数、SROCC 秩相关系数、RMSE 均方根误差、OR 背离率等参数来衡量。其中客观评分要先进行非线性压缩，再与主观评分做拟合。吻合度最高的方法为 puPSNR、puSSIM、HDR-VDP 、HDR-VQM ，这也是业内广泛接受并使用的方法。

9.2.2 无参考质量评价方法

尽管学术界对于无参考质量评价的方法很多，但是尚缺乏专门针对HDR进行优化，并得到业界广泛认可的方法。为完整起见，这里仅做简要介绍。随着 HDR 主观测试数据的逐渐丰富，NR 评价方法会逐步成熟。传统上，无参考（NR）质量评价有两大类方法：

A．有特定失真

即基于一些指标如模糊度、块效应、噪声、对比度等进行计算，常用的有 JND: just noticeable distortion[26],VAR: variance[27], LAP: laplacian[28], GRAD: gradien[28], FTM: frequency threshold metric[29], HPM: HP metric[30],Marziliano:Marziliano blurring metric[29]，KurtZhang:kurtosisbased metric[31]， KurtWav: kurtosis of wavelet coefficients [32], AutoCorr: auto correlation [28], RTBM: Riemannian tensor-based metric[33]；

由于 HDR 图像具有更大色域和更高亮度，指标评价上会考虑更多，如色彩丰富度、亮度等。

B．无特定失真，则有如下几类：

1）基于支持向量机

先用 SVM 进行失真类型识别，从而对特定失真类型进行 SVR 回归模型分析。如基于失真辨识的图像真实性和完整性评价(DIIVINE)算法[34]和盲/无参考图像空域质量评价(BRISQUE)[35] 算法。

2）基于概率模型的算法

这类方法首先建立图像特征与图像质量之间的统计概率模型，对待评价图像，提取特征后根据概率模型计算最大后验概率的图像质量来估计图像质量。如基于 DCT 统计信息盲图像完整性指数(BLIINDS)算法[36]和 ILNIQE 算法[37]。

3）基于码本的方法

这类方法通过聚类分析根据图像特征生成码本(词典)，并通过某种方式建立码本和图像质量之间的映射关系， 对待评价图像，提取特征后通过匹配码本来估计图像质量。如无参考图像质量评价码本表示(CORNIA) [38] 、BLISS[39]、dipIQ[40]、HOSA[41]等

4）基于神经网络的方法

与基于支持向量机的方法类似，这类方法先提取一定的图像变换域或空间特征，再基于已知质量数据训练一个神经网络回归分析模型，由图像特征预测图像质量。如 deepIQA[42] 和多任务端到端优化深度网络（MEON）[43] ，其中 MEON 的性能已经逼近主观评测方法，这也是未来研究的主要方向。

参考文献:

[1] ITU-R Rec.BT.500,Methodology for the subjective assessment of the quality of television pictures ,2012

[2] ITU-R Rec.BT.1788,Methodology for the subjective assessment of video quality in multimedia applications ,2007

[3] Hurst R N. The Future's So Bright, I Gotta Wear Shades: Test Patterns for HDR[C]// Technical Conference and Exhibition, Smpte. SMPTE, 2017.

[4] Hanhart P, Bernardo M V, Pereira M, et al. Benchmarking of objective quality metrics for HDR image quality assessment[J]. Eurasip Journal on Image & Video Processing, 2015, 2015(1):39.

[5] Jacob Munkberg, Petrik Clarberg, Jon Hasselgren, and Tomas Akenine-MÅNoller. High dynamic range texture compression for graphics hardware.ACM Trans. Graph., 25(3):698–706, 2006

[6] Tun.c Aydın, Rafa l Mantiuk, and Hans-Peter Seidel. Extending quality metrics to full luminance range images. In Electronic Imaging 2008, pages 68060B–68060B. International Society for Optics and Photonics, 2008.

[7] Z Wang, AC Bovik, A universal image quality index. IEEE Signal Process.Lett. 9(3), 81–84 (2002)

[8] Z Wang, AC Bovik, HR Sheikh, EP Simoncelli, Image quality assessment:from error visibility to structural similarity. IEEE Trans. Image Process.13(4), 600–612 (2004)

[9] Z Wang, EP Simoncelli, AC Bovik, in 37th Asilomar Conference on Signals, Systems and Computers. Multiscale structural similarity for image quality assessment, (2003)

[10] A Shnayderman, A Gusev, AM Eskicioglu, An SVD – based grayscale image quality measure for local and global assessment. IEEE Trans. Image Process. 15(2), 422–429 (2006)

[11] S Aja-Fernandéz, RSJ Estepar, C Alberola-Lopez, C-F Westin, in 28th Annual International Conference of the IEEE Engineering inMedicine and Biology Society. Image Quality Assessment based on Local Variance, (2006)

[12]HR Sheikh, AC Bovik, G de Veciana, An information fidelity criterion for image quality assessment using natural scene statistics. IEEE Trans. Image Process. 14(12), 2117–2128 (2005)

[13]HR Sheikh, AC Bovik, Image information and visual quality. IEEE Trans.Image Process. 15(2), 430–444 (2006)

[14] L Zhang, D Zhang, X Mou, D Zhang, FSIM: A feature similarity index for

image quality assessment. IEEE Trans. Image Process. 20(8), 2378–2386 (2011)
[15] Z Wang, Q Li, Information Content Weighting for Perceptual Image Quality Assessment. IEEE Trans. Image Process. 20(5), 1185–1198 (2011)
[16] XK Yang, WS Ling, ZK Lu, EP Ong, SS Yao, Just noticeable distortion model and its applications in video coding. Signal Process. Image Commun.20(7), 662–680 (2005)
[17] J Mannos, DJ Sakrison, The effects of a visual fidelity criterion of the encoding of images. IEEE Trans. Inf. Theory. 20(4), 525–536 (1974)
[18] T Mitsa, KL Varkur, in IEEE International Conference on Acoustics, Speech, and Signal Processing. Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms(1993)
[19] CIE, Colorimetry Official Recommendation of the International Commission on Illumination. CIE publication 15.2, CIE Central Bureau(1986)
[20] CIE, Industrial Colour-Difference Evaluation. CIE publication 116, CIE Central Bureau (1995)
[21] FJJ Clarke, R McDonald, B Rigg, Modification to the JPC79 Colour-difference Formula. J. Soc. Dye. Colour. 100(4), 128–132 (1984)
[22]M Luo, G Cui, B Rigg, The development of the CIE 2000 colour-difference formula: CIEDE2000. Color Res. Appl. 26(5), 340–350 (2001)
[23] R. Mantiuk, K. Kim, G. Rempe, et al. “HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions,” in ACM Transactions on Graphics, 2011, 30(4):1-14.
[24] M Narwaria, M Perreira Da Silva, P Le Callet, HDR-VQM: An objective quality measure for high dynamic range video. Signal Process. Image Commun. 35, 46–60 (2015)
[25] Banterle, Francesco. Advanced high dynamic range imaging : theory and practice[M]. A K Peters, 2011.
[26] XK Yang, WS Ling, ZK Lu, EP Ong, SS Yao, Just noticeable distortion model and its applications in video coding. Signal Process. Image Commun. 20(7), 662–680 (2005)
[27]SJ Erasmus, KCA Smith, An automatic focusing and astigmatism correction system for the SEM and CTEM. J. Microsc. 127(2), 185–199 (1982)
[28]CF Batten, Autofocusing and astigmatism correction in the scanning electron microscope. Master’s thesis. (University of Cambridge, UK, 2000)

[29] AV Murthy, LJ Karam, in 2nd International Workshop on Quality of Multimedia Experience. A MATLAB-based framework for image and video quality evaluation, (2010)
[30] D Shaked, I Tastl, in IEEE International Conference on Image Processing. Sharpness measure: towards automatic image enhancement, (2005)
[31] N Zhang, A Vladar, M Postek, B Larrabee, in Proceedings Section of Physical
and Engineering Sciences of American Statistical Society. A kurtosis-based statistitcal measure for two-dimensional processes and its application to image sharpness, (2003)
[32] R Ferzli, LJ Karam, J Caviedes, in 1st International Workshop on Video Processing and QualityMetrics for Consumer Electronics. A robust image Sharpness
[33] R Ferzli, LJ Karam, in 3rd International Workshop on Video Processing and QualityMetrics for Consumer Electronics. A no-reference objective sharpness metric using riemannian tensor, (2007)
[34] Moorthy A K, Bovik A C. Blind image quality assessment:from natural scene statistics to perceptual quality. IEEE
[35] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” IEEE Trans. Image Process., vol. 21,no. 12, pp. 4695–4708, Dec. 2012.
[36] Saad M A, Bovik A C, Charrier C. A DCT statistics-basedblind image quality index. IEEE Signal Processing Letters,2010, 17(6): 583-586
[37] L. Zhang, L. Zhang, and A. C. Bovik, “A feature-enriched completely blind image quality evaluator,” IEEE Trans. Image Process., vol. 24,no. 8, pp. 2579–2591, Aug. 2015.
[38] Ye P, Kumar J, Kang L, Doermann D. Unsupervised feature learning framework for no-reference image quality assessment. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI:IEEE, 2012. 1098-1105
[39] P. Ye, J. Kumar, and D. Doermann, “Beyond human opinion scores:Blind image quality assessment based on synthetic scores,” in Proc.IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 4241–4248.
[40] K. Ma, W. Liu, T. Liu, Z.Wang, and D. Tao, “dipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs,” IEEE Trans. Image Process., vol. 26, no. 8, pp. 3951–3964, Aug. 2017.

[41] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann, “Blind image quality assessment based on high order statistics aggregation,” IEEE Trans. Image Process., vol. 25, no. 9, pp. 4444–4457, Sep. 2016.
[42] S. Bosse, D. Maniry, K.-R. Mu¨ller, T. Wiegand, and W. Samek,“Deep neural networks for no-reference and full-reference image quality assessment,” CoRR, vol. abs/1612.01697, Dec. 2016.
[43] Ma K, Liu W, Zhang K, et al. End-to-end blind image quality assessment using deep neural networks[J]. IEEE Transactions on Image Processing, 2018, 27(3): 1202-1213.